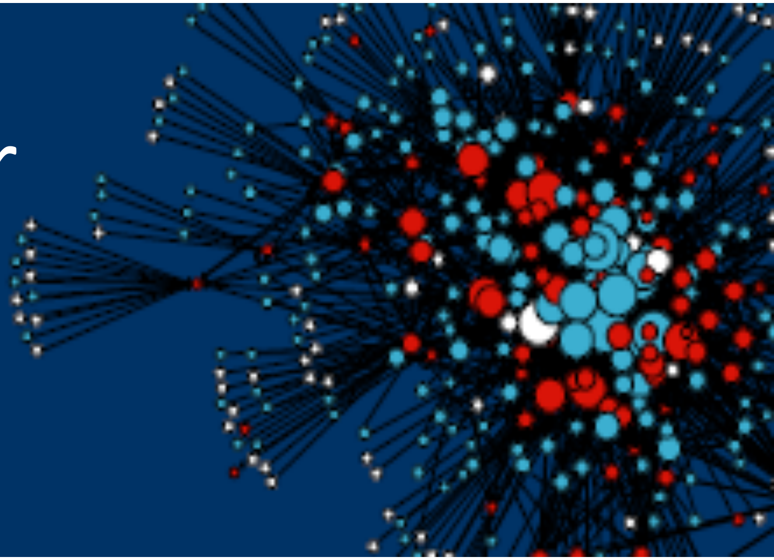
Complex Network Mining for Decision Making. (Specific Project)



Huzefa Rangwala, Ph.D.

BigData with ***Structure***: Large Graphs

social graph

social graph

follow-graph

consumer-products graph

user-movie ratings graph

DNA interaction graph

WWW link graph

Social Networks

".. defined between persons or groups of persons with some pattern of interactions or connections amongst them."

EXAMPLES:

- **Friend-to-friend Networks**
- **Actor-to-actor Networks**
- **Email Networks**
- **Blogger Networks**
- **Reply Networks**

Explicit Relationship Networks

Explicit relationships (friends, enemies, followers, professional colleagues) are defined between the entities (nodes/people) within the network.

Nodes can belong to multiple communities or have different properties.
Edges can be labeled, or have weights or just binary.

DBLP Co-participation Networks

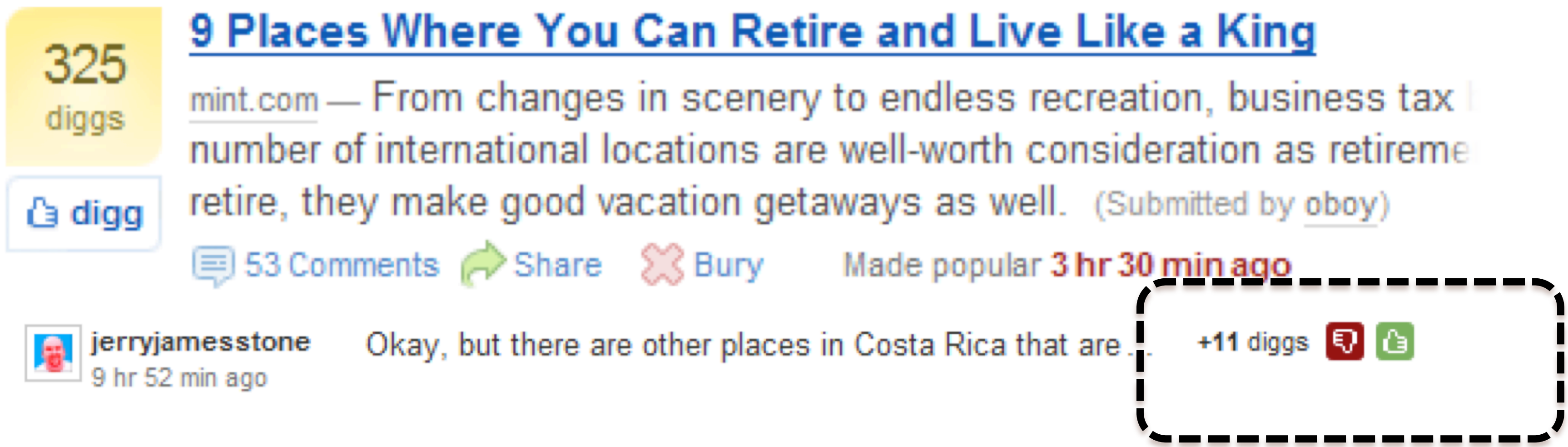
The DBLP server provides bibliographic information on major computer science journals and proceedings.

Several papers analyze the co-authorship network as well as the citation network derived from DBLP database.

Social Bookmarking websites

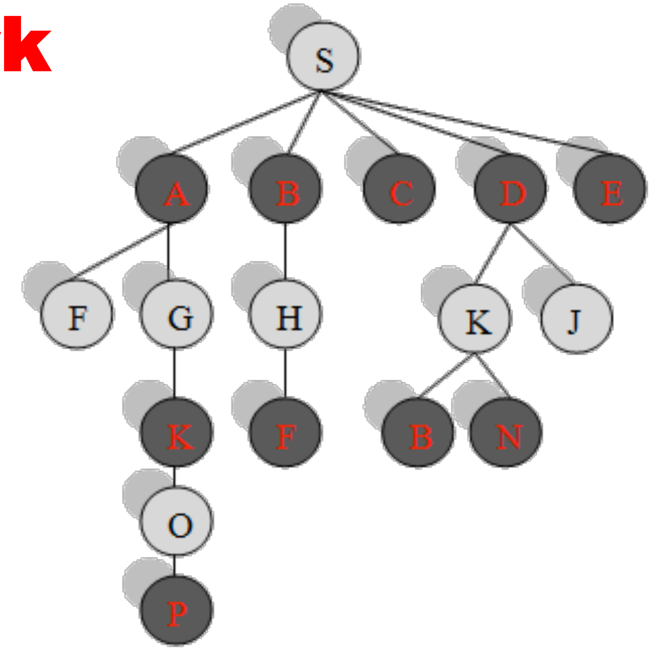
"A Web-based service where users can create and store links. It is an increasingly popular way to locate, classify, rank, and share internet resources" – FDLP

Digg Definitions

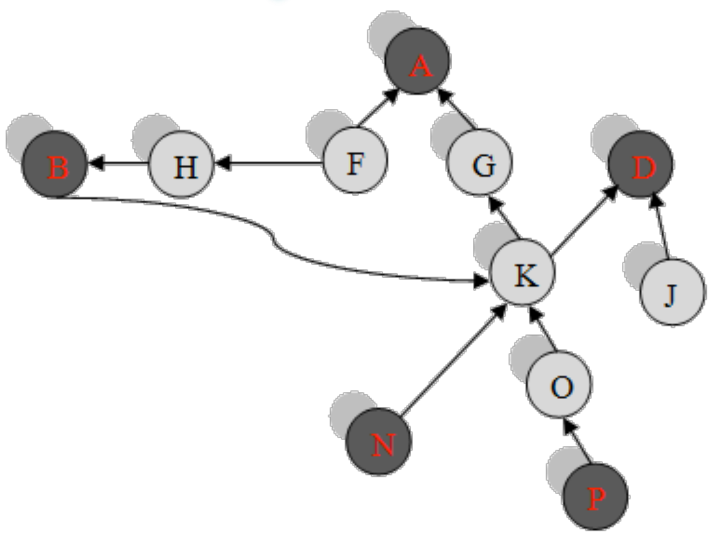


- **story** – a social bookmark
- **user** – contributor and/or commenter
- **digg** – positive rating
- **bury** – negative rating
- **category** – main topics
  - Sports, Business, Science, etc.
- **topic** – sub topics
  - Linux, Elections, Golf, etc.

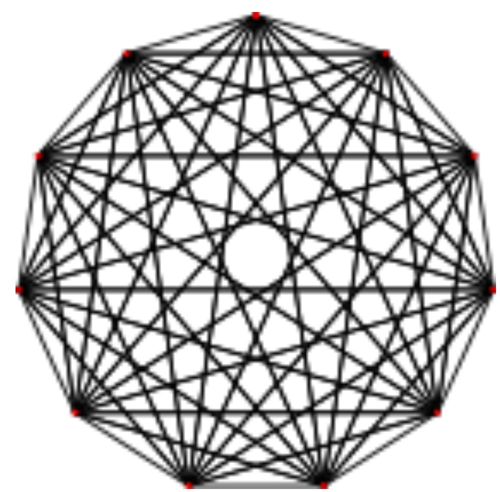
Digg Implicit Network



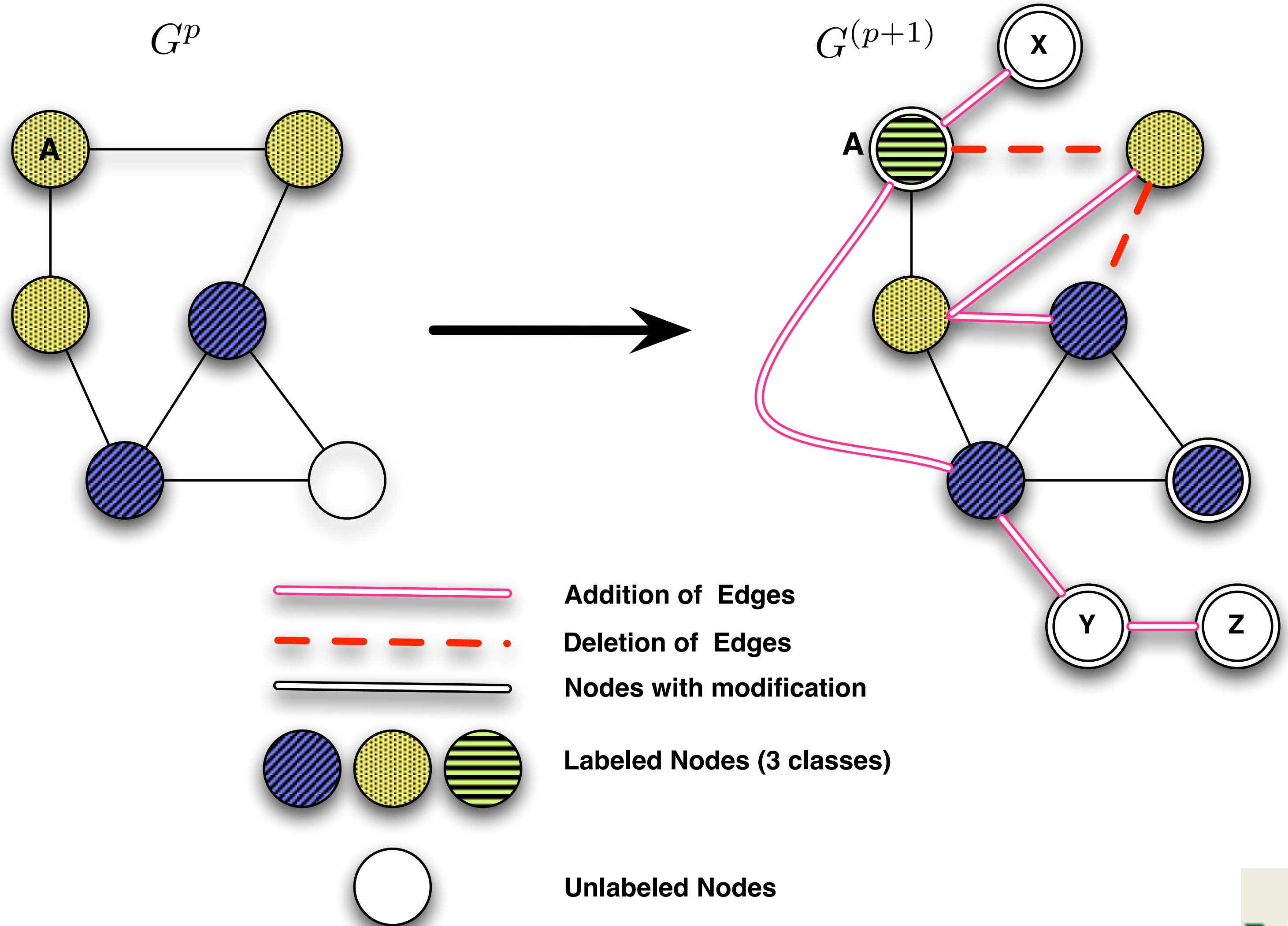
Reply Network



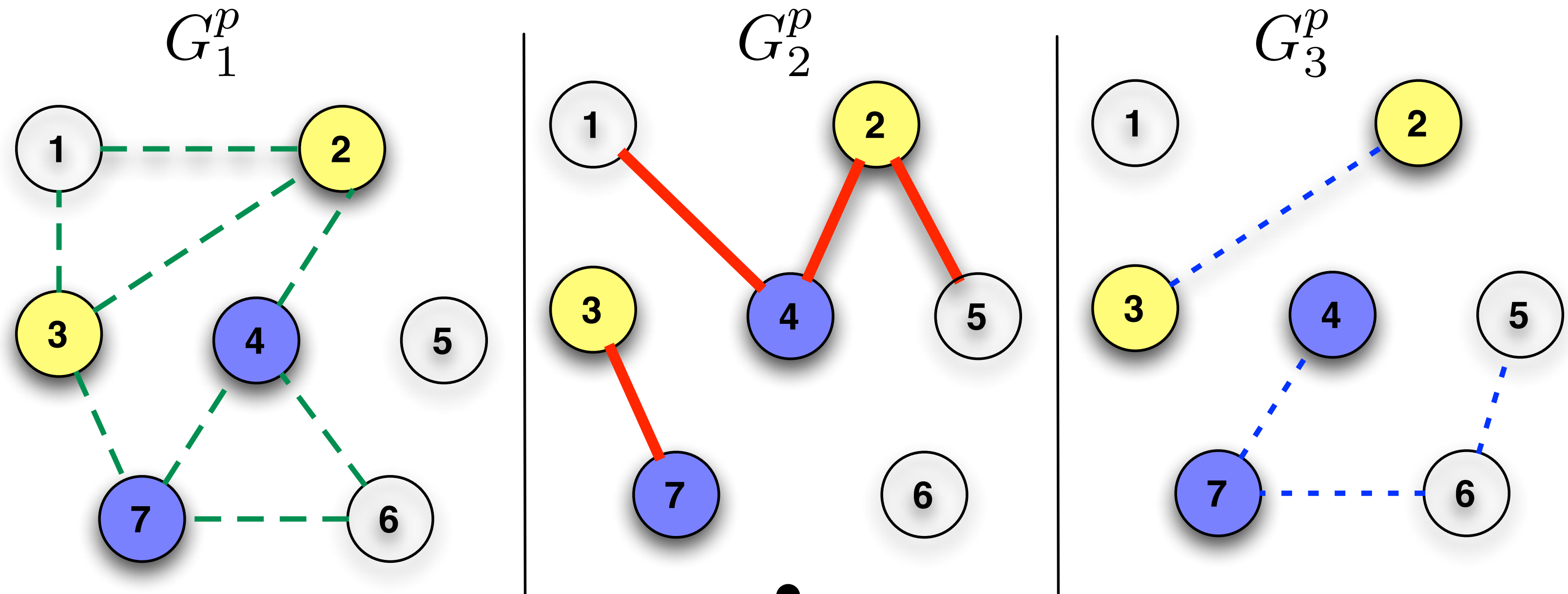
Co-participation Network



Defining Complex Networks



Complex Multi-Relational

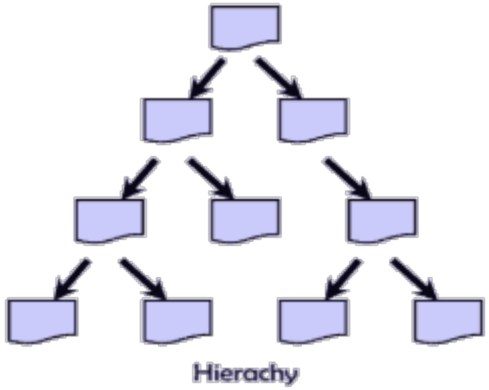


Infer Latent Edges using EM

$$\Pi^p = \begin{bmatrix} \pi^p_{1,1} & \cdots & \cdots & \pi^p_{1,7} \\ \cdots & \cdots & \cdots & \cdots \\ \pi^p_{7,1} & \cdots & \cdots & \pi^p_{7,7} \end{bmatrix}_{7\times 7}$$

Output can be ``Structured''

- Not 0/1 classification or regression
  - But relationship between output classes/ variables.
- Examples:
  - Multi-labeled
  - Hierarchical
  - Partially Labeled



	f1	f2	f3	f4
p1	?	1	0	0
p2	0	1	?	0
p3	1	?	0	?
p4	0	?	1	0
p5	?	0	0	1
p6	0	?	1	0

- Other Challenge: Several Thousands of Classes

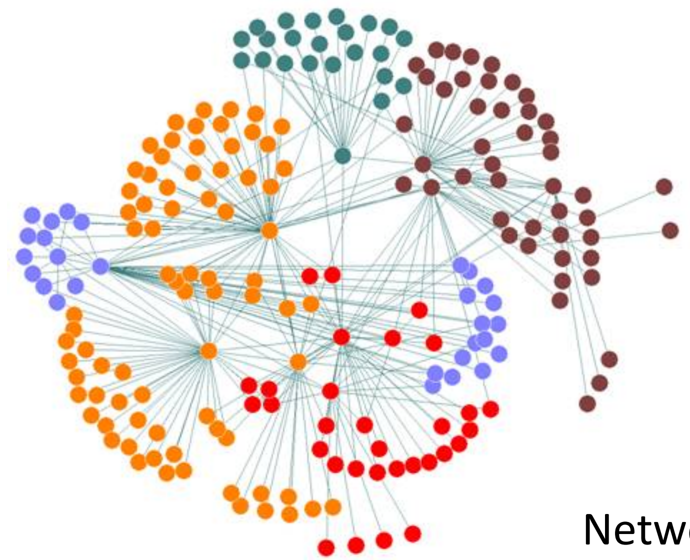
Determining a Node (Collective Classification)

Input: A graph G = (V,E) with given percentage of labeled nodes for training, node features for all the nodes

Output: Predicted labels of the test nodes

Model:

- Relational features and node features are used for training local classifier using labeled nodes
- Test nodes labels are initialized with labels predicted by local classifier using node attributes
- Inference through iterative classification of test nodes until convergence criterion reached



Network of researchers

?

Collective Classification

Multi-labeled collective classification (Kong et. al. 2011)

- Assume “K” possible labels.
- Initialization: Train “K” one-versus-rest classification models for the different labels.
  - Use only train nodes.
  - Features: Attributes, Self-Label Features (i.e., other labels)
- Repeat
  - Predict labels for test nodes.
  - Retrain ``baseline” models.
    - Features: Attributes, Self-Label Features, Cross-labeled features (from neighboring nodes).
- Until convergence (Labels do not change).

Our Approach

Multi-labeled collective classification using ranked neighbors (Saha et. al. 2012, 2013)

Intuition:

- Are there influencing neighbors ?
- Are some of the more important ?
- Can we use a ranking based list ?
- Can we speed up the computation by removal of edges that do not convey any information ? – sparsification?
- Active-learning approach.

For Baseline Model Learning

Obtain,

$$f\left(x \mid w\right) \sim y$$

Objective Function,

$$\min_w \Sigma_{i=1}^{N} \mathcal{L}\left(x_i, y_i, w\right) + \lambda \mathcal{R}\left(w\right)$$

Loss Function (Hinge Loss, Least Squares, Logistic Loss)

Regularization Term (usually a norm of w)

Can we couple models across different time periods ?

Obtain,

$$f_t\left(x \mid w\right) \sim y_t \qquad \forall t \in \{1, 2, \ldots, T\}$$

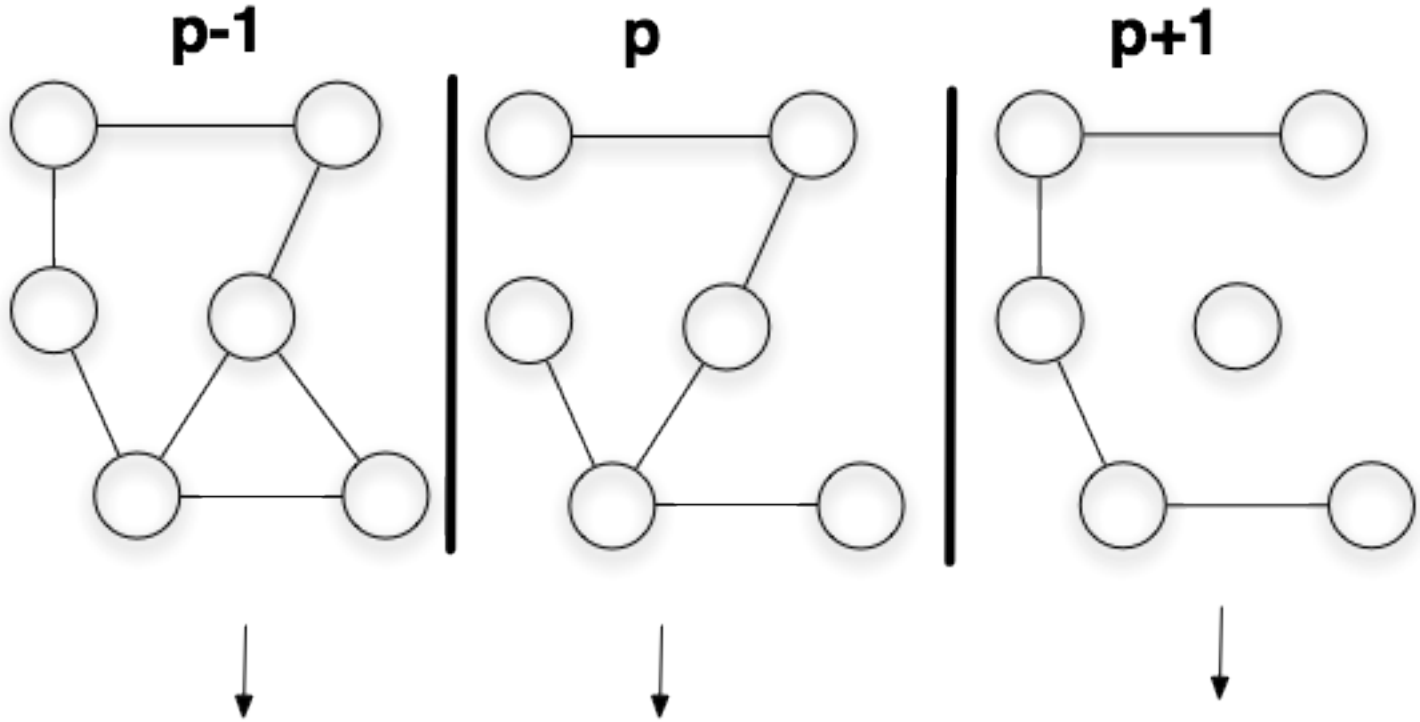
Objective Function,

$$\min_{W} \Sigma_{t=1}^{T} \Sigma_{i=1}^{n_t} \mathcal{L}\left(f\left(x_i \mid W_t\right), y_i\right) + \lambda \mathcal{R}\left(\{W_t\}_{t=1}^{T}\right)$$

Loss (sums the misclassification error over all the examples from all the tasks)

Regularization term jointly regularizes the model parameters of all the tasks.

Jointly/Iteratively Learn Model Parameters



$$W^{(p-1)} = [w_1^{(p-1)}, \ldots, w_k^{(p-1)}] \qquad W^p \qquad W^{(p+1)}$$

Different Regularization Penalties

Joint Feature Selection (Assume shared Features)

$$\mathcal{R}(\mathcal{W}) = \|W\|_{2,1}$$

Difference in two periods

$$\mathcal{R}(W) = \sum_{(p,q)\in\mathcal{E}} \|W_p - W_q\|_2^2$$

Advantages

- Better generalization of jointly trained parameters
  - Relationship across epochs.
- Need fewer labeled examples.
  - Scarcity in training data supply.

BIG Data presents BIG problems.

- Big Parameters. Extreme classes. Large Dimensions.
- Need iterative/concurrent formulations for standard optimization techniques.
- MPI/Hadoop/Distributed version.
- Need local network and time variant estimation properties of the algorithms.
- Other Questions?
  - Early Time Classification.
  - Human in the loop (Active Learning Approaches).
  - Detection of Dynamic Network Patterns.
    - No standard definitions.